Density Functional Theory and Biomolecules: A Study of Glycine, Alanine, and Their Oligopeptides

R. Kaschner* and D. Hohl†

Institut für Festkörperforschung, Forschungszentrum Jülich, D-52425 Jülich, Germany

Received: February 2, 1998; In Final Form: April 6, 1998

We present density functional (DF) calculations, using a pseudopotential scheme and plane waves as basis functions, for isolated molecules of the amino acids glycine and alanine, for small oligopeptides composed of glycine and alanine, and for periodic (infinite) polyalanine helices. We calculate relative energies and geometries for the low-lying isomers of glycine and alanine and for a variety of oligopeptide geometries using various DF formulations for electron exchange and correlation (LDA, PBE, BLYP, BP). Comparison is made with other theories and experiment where possible. The free molecule equilibrium geometries agree well with the limited experimental data and with post-Hartree−Fock (post-HF) calculations. The inclusion of gradient-corrected (nonlocal) functionals is essential when hydrogen bonds play a role in determining relative energies. This is especially true for hydrogen bonds of the type N···H−O, which appear in two isomers of glycine and alanine. We obtain the most reliable results with BLYP, but the best compromise, with a considerably smaller cutoff energy, is PBE. For the polypeptides we find that the peptide bonds in the equilibrium geometries are planar to high accuracy, with dihedral angles deviating from planarity by up to 15°. The relative energies of the low-lying isomers of alanine dipeptide agree very well with post-HF calculations. The equilibrium structure of the polyalanine α helix is very well reproduced by our calculations.

1. Introduction

The theoretical treatment of biologically active molecules is a mature and still exceptionally active and challenging research area. The main difficulties are *complexity* and the *high level of precision* required. Biomolecules often contain hundreds or thousands of atoms and must therefore be approached with approximate methods. A prominent example is protein folding,[1−4] where one seeks the biologically active ("native") minimum free energy structure for heteropolymers with several hundred atoms, a virtually insoluble optimization problem in a very high-dimensional space.[5] The native structures are often stabilized over other, inactive, ones by only 5−10 kcal/mol. Atomic-level simulations[1] are commonly restricted to short trajectories and use empirical interatomic forces that are often not transferable to systems with a different chemical bonding character. Ab initio (AI) methods based on solving the quantum mechanical interacting electron−ion problem with no adjustable parameters promise better accuracy and no transferability problem. In practice, because of computational demands and fundamental limitations (scaling), traditional AI methods such as Hartree−Fock and correlated wave function approaches are confined to small molecules and providing a limited database for fitting empirical potential parameters.[6] AI methods based on Hohenberg−Kohn−Sham density functional theory[7,8] in combination with faster (parallel) computers have greatly expanded the range of directly accessible systems. Biopolymers with 200 and more atoms are tractable with these methods (see, e.g., ref 9), at what appears to be an acceptable level of precision. The purpose of this contribution is to evaluate in detail both the accuracy and practical applicability of density functional (DF) methods to amino acids and small peptides. In this respect, our study is similar in scope to the work of Montanari and Jones for hydrocarbons.[10] We use a pseudopotential scheme with plane waves as basis functions (section 2). The smallest system we study is the glycine molecule (10 atoms); the largest is polyalanine in the α helix conformation (70 atoms in the supercell).

An array of AI studies for the glycine molecule is already available,[11−14] and it is well established that this small molecule poses a very formidable challenge for computational theory. Intramolecular hydrogen bonds require the inclusion of electron correlation at high levels of theory. Basis set convergence is often poor, so that it is not surprising that even the most sophisticated calculations agree on relative energies to only about 1 kcal/mol. Within density functional theory (DFT), the choice of the electron exchange-correlation functional plays a crucial role.

In section 3.1 we consider all relevant glycine isomers (including the zwitterion) in the local density approximation (LDA) and in three common gradient-corrected functionals (BP,[15,16] BLYP,[15,17] and PBE;[18] see section 2).

A similar study of structure and energetics in the next more complex amino acid alanine (see section 3.2) is the first systematic DFT investigation for this molecule, and the results are qualitatively and/or quantitatively comparable to those for glycine. Finally, in section 3.3, we investigate the pertinent issues of structure and energetics when amino acids polymerize to form peptides. We present calculations for the di- and triglycine peptides, for alanine dipeptide, and for an infinite chain of the polyalanine helix in several conformations.

2. Computational Methodology

The (Hohenberg−Kohn−Sham) DFT calculations described in this article were performed with the computer program

* Corresponding author. E-mail: r.kaschner@fz-juelich.de.
† Permanent address: Shell E&P Technology Co., Geophysics Department, Houston, TX 77001-0481. E-mail: dhohl@shell.com.

S1089-5639(98)00975-X CCC: $15.00
Published on Web 06/09/1998

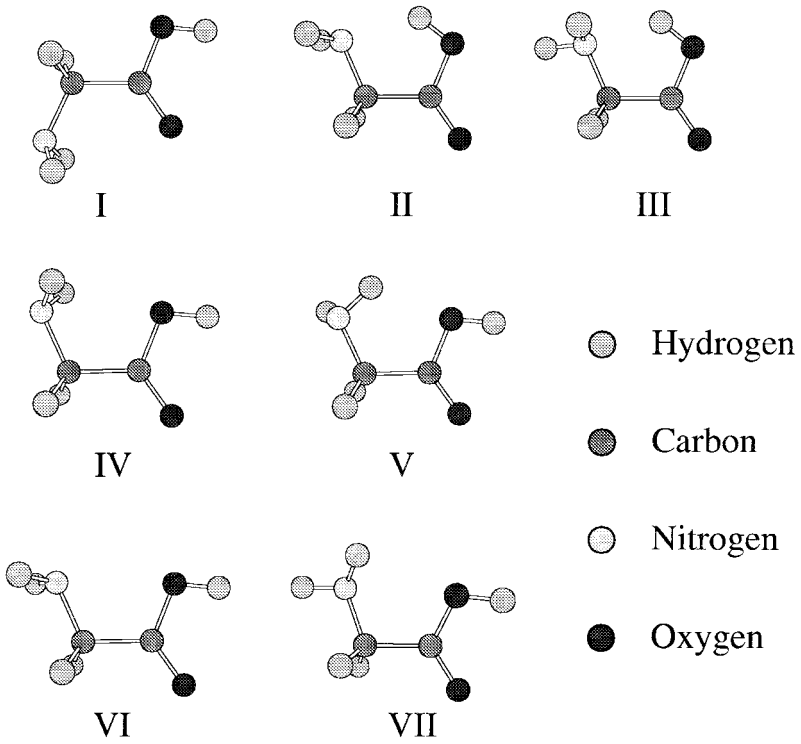

Figure 1. Structure of the seven (**I**−**VII**) most stable glycine isomers H_2N-CH_2-COOH. Isomers **III** and **II**, isomers **V** and **IV**, and isomers **VII** and **VI** are slightly symmetry-distorted versions of each other.

CPMD.[19] It is based on the Car-Parrinello method[8] and uses plane-wave expansions for the electronic Kohn−Sham eigenfunctions and electron density in conjunction with a pseudopotential formulation for the electron−ion interaction (for reviews see ref 20). The plane-wave basis implies that only one parameter (the cutoff energy E_{cut}) controls the basis set convergence. Only the valence electrons are included explicitly in the electronic structure calculations.

We have used nonlocal norm-conserving Troullier−Martins type[21] pseudopotentials with matching radii of 0.8 au for hydrogen (local only), 1.1 au for carbon, and 1.3 au for nitrogen and oxygen. The pseudopotentials were generated separately for LDA and the three types of gradient corrections: (i) BP (i.e., the expression of Becke for exchange[15] and of Perdew for correlation[16]), (ii) BLYP, the identical expression for exchange[15] and of Lee, Yang, and Parr for correlation,[17] and (iii) PBE, the new simplified version of the generalized gradient approximation (GGA), by Perdew, Burke, and Ernzerhof.[18] The plane−wave cutoff for the electronic wave functions was 60 Ryd for LDA, 70 Ryd for PBE, 90 Ryd for BP, and 130 Ryd for BLYP. The necessity for the high E_{cut} value for BLYP originates from large oscillations in the BLYP pseudopotentials for oxygen and nitrogen.

The plane−wave expansion implicitly requires the consideration of the isolated molecules in supercells which are periodically repeated in space. The supercells were 23 × 23 × 23 au for glycine and alanine, 30 × 30 × 30 au for the diglycine molecule and the alanine dipeptides, and 36 × 36 × 36 au for the triglycine molecule. The supercell of the polyalanine helix calculation is described in section 3.3.3. These large supercells in real space imply the existence of a small Brillouin zone. Hence, for the integration over the reciprocal space, the Γ point approximation was used.

Most of the calculations were performed on a parallel computer (Cray T3E). The electronic wave functions were optimized using the DIIS method (direct inversion of the iterative subspace);[22] for geometry optimization, the nuclei were relaxed with a quasi-Newton method using the Broyden−Fletcher−Goldfarb−Shano (BFGS) procedure.[23] These schemes work very efficiently and allow the geometry optimization of, for example, one glycine or alanine molecule with E_{cut} =150 Ryd with 64 processors on the T3E computer in 15 dedicated minutes. The corresponding (converged) electronic structure calculation for one geometry takes about 1−2 min. For E_{cut} = 70 Ryd the computing time is smaller by a factor of about 2.5.

3. Results

3.1. Glycine. Calculating relative energies reliably for the most stable isomers of the glycine molecule (Figure 1) is a very difficult task for *any* parameter-free method. The main problem is the determination of the energy difference between isomer **I** and isomers **II** and **III**, experimentally known to be +1 ± 0.5 kcal/mol. In this contribution, we follow the convention of Barone et al.[12] for numbering the isomers.

With the exception of the zwitterion (see below) the various glycine structures are rotational isomers around the C−N, C−C, and C−O single bonds. All of the seven most stable isomers considered here have *intra*molecular hydrogen bonds of the types N−H···O (**I**, **IV**, **V**, **VII**), O−H···N (**II**, **III**) or O−H···O (**I**, **IV**, **V**),[11,12] see Figure 1. Weak closed-shell interactions of this type are very sensitive to the level of electron correlation and the basis set employed.[13]

We have optimized the geometries of isomers **I** to **VII** on various levels of sophistication for electron exchange and correlation within DFT, and the results can be found in Tables 1 and 2. The corresponding numbers from selected previous studies, along with the experimental values,[24] are included for comparison.

There is some debate in the literature about the character of some of the structures not being genuine local minima but saddle points on the energy surface (e.g., in ref 12). We have used explicit total energy calculations in the vicinity of the minima

TABLE 1: Structural Parameters of the Optimized Structure of Glycine I, Compared with Experiment, HF/CISD Calculation, and a Previous LCAO LDA Calculation[a]

	exptl[b]	HF/CISD[c]	LDA[d]	LDA	PBE	MM
N−H	(1.001)	1.011	1.028	1.024	1.023	1.026
N−C	1.467	1.445	1.439	1.437	1.448	1.480
C−H	(1.081)	1.090	1.109	1.109	1.103	1.096
C−C	1.526	1.518	1.510	1.499	1.523	1.472
C−O	1.355	1.344	1.348	1.363	1.360	1.285
O−H	(0.966)	0.962	0.988	0.983	0.982	0.974
C═O	1.205	1.204	1.218	1.212	1.216	1.318
H−N−H	(110.3)	105.3	105.0	105.6	105.5	100.9
H−N−C	(113.3)	109.5	109.0	109.9	109.9	106.2
N−C−C	112.1	115.2	114.8	114.7	115.4	116.4
H−C−H	(107.0)	106.2	104.8	104.6	105.2	106.9
H−C−C		107.4	107.6	107.9	107.6	108.7
C−C−O	111.6	111.5	112.1	112.4	111.6	124.0
C−O−H	(112.3)	107.1	105.6	105.2	106.2	101.8
C−C═O	125.1	125.6	124.9	125.1	125.3	126.2

[a] Angles in degrees, distances in angstroms. The uncertainties in the experimental numbers are large (see text); and values in parentheses were assumed in ref 24, not measured. The last three columns are the present results: LDA-DFT, PBE-DFT, and molecular mechanics with the Merck molecular (MM) force field.[26] [b] Reference 24. [c] Reference 11. [d] Reference 12.

TABLE 2: Relative Energies for the Glycine Isomers in kcal/mol, Compared with Results from Hartree−Fock/CISD (Reference 11) and Hartree−Fock/MP Calculations (Reference 14). See Also Reference 12

isomer	HF/CISD	HF/MP	LDA	BP	BLYP	PBE
I	0.00	0.00	0.00	0.00	0.00	0.00
II	1.90	0.56	−3.64	−1.37	−0.43	−2.01
III	1.77	0.49	−3.60	−1.30	−0.36	−1.91
IV	1.70	1.59	1.42	1.47	1.45	1.40
V	1.71	1.60	1.49	1.61	1.56	1.58
VI	5.80	5.01	4.52	4.61	4.50	4.65
VII	2.71	2.51	2.51	2.69	2.57	2.60

(by small displacements of the atomic coordinates) to test for such behavior. We obtain isomer **VI** as a saddle point on the energy surface, in agreement with ref 12. Isomers **II** and **IV**, which are also saddle points in most post-Hartree−Fock (post-HF) calculations,[12] are true minima in our calculations. The more symmetric (C_s) structures **II** and **IV** are *more stable* in DFT than the distorted geometries **III** and **V**. Note, however, that the absolute energy differences are very small for all computing methods, see Table 2.

The structures that we obtain agree very well with post-HF results [e.g., the HF/CISD (configuration interaction with single and double excitations) calculations[11] (see Table 1 for isomer **I**)]. This is true already within LDA and the structures generally show little sensitivity to the level of electron exchange and correlation.[25] Within the PBE gradient correction[18] the geometrical parameters, in particular the angles, agree slightly better with the HF/CISD calculation than with LDA. The agreement is within about 2% for the bond lengths (in most cases within 1%) and within 1° for the angles. Few experimental studies of structural parameters for the gas-phase molecule exist,[24] and the uncertainties in the resulting bond distances and angles can be larger than for the most accurate theoretical results. One example is the angle C−O−H in the carboxyl group assumed in ref 24 to be 112.3°, but consistently between 106 and 107° in all high-level theoretical studies (see Table 1).

Overall, our plane−wave pseudopotential-based DFT at the PBE gradient-corrected level describes the structure of the glycine molecule very satisfactorily, at a fraction of the computational cost for post-HF methods and without compromising convergence with respect to basis set size. Basis set issues are also the most likely source of differences between our own DFT calculations and ref 12.

To assess the reliability of the best available empirical force-field methods, we have also studied glycine with a molecular mechanics method. The Merck molecular force field (MMFF)[26] is one of the most sophisticated force fields available for organic molecules today, comprising bond stretch, van der Waals and electrostatic (two-body) terms, angle bending (three-body) terms and torsional (four-body) forces. The parameters are determined by fitting to a very large experimental and AI database.

MMFF indeed successfully locates a stable minimum for glycine **I**, lower in energy than all other isomers. The structure of glycine **I** with MMFF is also shown in Table 1. The agreement with the experimental, HF/CISD and our structure data in Table 1 is only satisfactory. The MMFF bond distances deviate by up to 10% from the other values; the angles deviate by up to 13°. Unusual chemical bonding features such as intramolecular hydrogen bonds and partial double bonds influence the structure of this small molecule. Empirical force fields, even when fitted to a very extensive database like MMFF, will not properly describe the ensuing sometimes subtle geometry changes (e.g., the C−O and C═O bond lengths and the bond angles). Nevertheless, given the negligible computational effort for MM-type calculations, MMFF seems to perform well for a complex structure prediction such as glycine **I**.

The problems inherent in force-field methods such as MMFF become more apparent when one attempts to study the less stable isomers **II**−**VII**. We find that only glycine **IV** has a stable local minimum on the MMFF energy surface, 5.5 kcal/mol above the ground state. Glycine **III**, **VI**, and **VII** transform without barrier into a nonplanar structure, which lies only 0.45 kcal/mol above glycine **I** and is not seen in ab initio studies, with a dihedral angle O−C−C−N of ∼80°. Glycine **V** transforms without barrier into glycine **I**, and glycine **II**, when kept in C_s symmetry, is more than 9 kcal/mol less stable than glycine **I**.

The relative energies of the glycine isomers show more interesting trends than the structures. These energies, along with results from previous work, are given in Table 2. In addition to the more recent PBE formulation of electron exchange and correlation, we also list the results for BP and BLYP.

Within a few 0.1 kcal/mol, the results for all density functionals agree with each other and with HF/MP[14] and HF/CISD[11] calculations. The notable exceptions are isomers **II** and **III**, a slightly distorted version of **II**. These isomers differ from all others in that they display O−H···N intramolecular hydrogen bonds (see Figure 1) which are energetically more important than the other types of hydrogen bridges mentioned above; cf. refs 11 and 12. The LDA overestimates this bond strength greatly, which corresponds to a lower (more negative) binding and relative energy and to a smaller hydrogen bond length (d_{N-H} = 1.71 Å; for comparison, with HF/CISD d_{N-H} =1.94 Å[13]). For PBE and BP this overestimate is much less (d_{N-H} =1.79 and 1.83 Å, respectively), and for BLYP even less (d_{N-H} =1.90 Å), see the energies in Table 2. It appears plausible that overestimating the strength of this hydrogen bond and the stability of isomers **II**/**III** relative to **I** are correlated.

Overestimating the strength of weakly interacting closed-shell fragments is a well-known deficiency of the LDA,[25] and gradient corrected schemes provide, if at all,[10] only partial improvement. In our study *no* density functional yields the correct energetic ordering for isomers **I** and **II**/**III**, although the gradient corrected schemes perform markedly better than LDA. BLYP performs best for the **I** vs **II**/**III** energy difference, and BP yields the

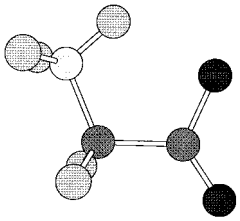

Figure 2. Structure of the glycine zwitterion $H_3N^+-CH_2-COO^-$. Labeling of the atoms as in Figure 1.

TABLE 3: Relative Energies for Some Alanine Isomers in kcal/mol Compared with Results of a Post-HF (MP2,MP3,MP4) Study (Reference 28)

isomer	HF/MP	LDA	BP	BLYP	PBE
I	0.00	0.00	0.00	0.00	0.00
II	−0.04...1.10	−4.26	−1.79	−0.72	−2.50
III	0.01...1.45	−4.18	−1.72	−0.65	−2.45
IV	1.03...1.51	1.42	1.22	1.09	1.32
V	1.08...1.51	1.08	0.92	0.80	0.98
IX	2.18...2.55	2.31	2.44	2.21	2.49

best agreement for the other isomers. It is important to realize that the absolute energy differences involved are minute. Not surprisingly, another recent DFT study,[13] using a Gaussian basis, finds isomer I more stable than **II**/**III** using the same BLYP functional that we have employed here; see also the BLYP relative energies in ref 12, which lie between our BLYP values and those of ref 13. The convergence of the basis sets in other studies, and the use of a pseudopotential in our own work can lead to apparent discrepancies of 0.5−0.6 kcal/mol (∼0.02 eV/atom) between otherwise identical calculations. It appears that for glycine and other similar biomolecules, all DFT results within this (systematic) uncertainty are equally reliable for the purpose of predicting energetics.

We have also investigated the gas-phase glycine zwitterion (see Figure 2). The zwitterionic form of glycine exists in the crystal and in aqueous solution. With all our choices of the electron exchange-correlation functional the gas-phase zwitterion transforms without barrier into the neutral isomer **II** (see Figure 1). This is consistent with good post-HF studies of the same system.[12]

3.2. Alanine. Alanine differs from glycine only through a methyl group on the α-carbon atom. Our results show that the introduction of this additional rotational degree of freedom has virtually no influence on the structure and relative energetics of the dominant isomers.

Using an analogous numbering scheme and notation as in glycine, we list the relative energies of six alanine isomers in Table 3. Alanine is already a sufficiently complex molecule that very few quantum chemical studies exist (and no reliable experimental measurements are available). Table 3 also lists the range of relative energies obtained in the post-HF study.[28] Like in glycine, our relative energies for the alanine isomers agree well with that study, with the notable exception of isomers **II** and **III**, especially at the LDA level. The underlying reason is the same as for the glycine molecule, an intramolecular hydrogen bond of the type O−H···N that appears to be vastly overbound in LDA and is better described but still too stable using the BP, BLYP, and PBE functionals.

The minimum-energy structures that we derive also agree well with ref 28 for all isomers considered, with deviations that are about as large as for glycine (see Table 1). We consistently derive C−H distances of 1.10 Å in the methyl group and a C−C bond distance of 1.51 Å between the α- and methyl carbon atoms. The C−C sp^3 single bond is always in the *staggered* conformation. We have again studied the zwitterion, and found that it transforms to the neutral isomer **II** without crossing a barrier as in the case of glycine, irrespective of the exchange−correlation functional used.

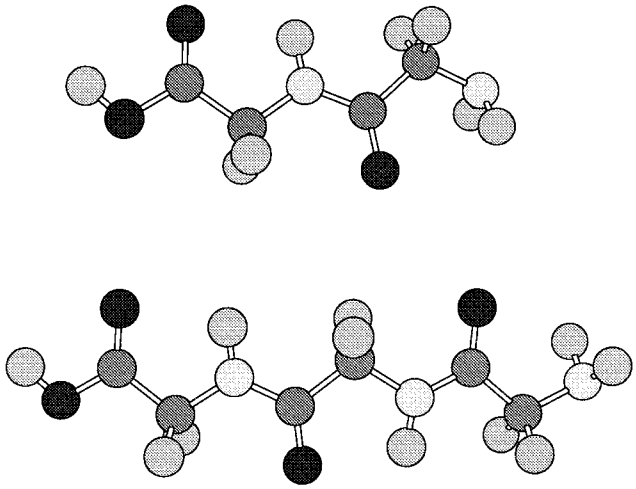

Figure 3. Di- and triglycine (stretched β form). Labeling of the atoms as in Figures 1 and 2.

3.3. Small Polypeptides. We have studied the small molecules glycine and alanine in some detail in the preceding sections primarily as a gauge of precision and feasibility for DF methods. DFT has the advantage over other AI methods that much larger systems can still be studied with relative ease. In the following section we describe calculations for small peptides (diglycine and triglycine, alanine dipeptides, the polyalanine helix). A correct description of peptide bond properties is essential for a reliable description of these systems. Again, our primary goal here is the computation of structures and relative energies, and an evaluation of the accuracy and efficiency with which DFT can perform such calculations routinely.

Systematic studies of peptides with more traditional quantum chemical methods such as HF and correlated wave function methods are not available because of the prohibitive computational effort. Ample experimental data exist from many sources, and we shall compare our results with these where possible.

All calculations presented below were performed using the PBE functional for electron exchange and correlation, with E_{cut} = 70 Ryd. We find that this functional represents a good compromise between computational effort and precision in biomolecular systems with hydrogen bonding.

3.3.1. Diglycine and Triglycine. The simplest molecules with one and two peptide bonds−(HN)−(C═O)− are the stretched forms of the glycine di- and trimer, di- (DiGly) and triglycine (TriGly), in the β conformation, see Figure 3. The geometry optimization of a triglycine molecule in the large supercell (see section 2) takes about 2 dedicated hours on a Cray T3E using 64 processors, with about 5 min for the (converged) electronic structure calculation per geometry. For the fully optimized geometries we give the structural parameters around the peptide bonds in Table 4. All other angles and bond lengths agree to within 0.002 Å and 1° with the corresponding numbers in Glycine **I** (see Table 1).

Table 4 shows that the geometric parameters of all three peptide bonds agree very well with each other and with the experimental data. Compared with the corresponding bond lengths in Table 1, only the C−N distance is significantly different in the peptide bond: It is about 0.1 Å shorter than the C−N distance in glycine (or than the C_α−N distance in di- and

TABLE 4: Peptide Bond Geometric Parameters for the Equilibrium Geometries of Diglycine and Triglycine in Their β (stretched) Conformation[a]

	DiGly	TriGly I	TriGly II	exptl
$C_\alpha-N$	1.440	1.442	1.441	1.46
N−H	1.021	1.022	1.025	
C−N	1.360	1.353	1.359	1.32
C═O	1.234	1.236	1.235	1.24
$C-C_\alpha$	1.534	1.524	1.534	1.53
$C_\alpha-N-C$	121.6	121.6	121.5	
H−N−C	122.4	122.5	122.1	
N−C═O	122.1	122.5	122.4	
$N-C-C_\alpha$	115.6	115.6	115.5	
$C-C_\alpha-N$	108.8	108.4	108.1	

[a] TriGly **I** and TriGly **II** designate the two peptide bonds in triglycine. Distances are in angstroms, angles in degrees. The experimental data are from ref 4.

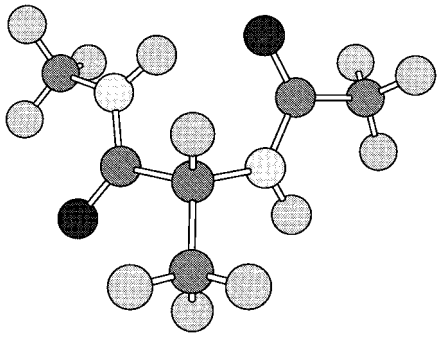

Figure 4. The most stable conformer (C_7^{eq}) of alanine dipeptide. Labeling of the atoms as in Figures 1−3.

triglycine). This is explained by the partial double-bond character of C−N in the peptide bond.[4] The corresponding partial (incomplete) C═O double-bond character in the peptide group slightly increases the C═O bond length (by 0.02 Å).

To check the planarity of the peptide bonds,[4] we initiated structural relaxations from distorted nonplanar arrangement geometries. We find that the peptide bonds consistently relax to planarity with high accuracy in the final equilibrium geometries. The dihedral angles between the atoms of the peptide bonds deviate from planar (0° or 180°) by up to 5° in diglycine and up to 10° in triglycine. Hence, conformations with *slightly* nonplanar peptide bonds can be more stable than the (fully) planar geometry. The energy surfaces near the planar minima are very shallow.

3.3.2 Alanine Dipeptide. A more realistic reference system for a polypeptide is the alanine dipeptide: a shortened tri-alanine molecule with two peptide bonds, saturated with CH_3 groups at the ends next to the peptide bonds, see Figure 4. This system can be described by the dihedral angles ϕ and ψ used in the Ramachandran maps of peptides.[4] The alanine dipeptide was recently investigated using the MP2 and B3LYP methods,[29] see also ref 30 for earlier HF-based calculations. The present study is the first systematic investigation of this formidable molecule with DFT and pseudopotentials.

We have optimized the geometries of five conformers of this system and compare the resulting energies and angles ϕ and ψ with the calculations of ref 29 in Table 5. The relative energies agree very well and the angles ϕ and ψ match qualitatively, a reflection of the fact that the total energies of these molecules depend only weakly on the dihedral angles (see above). The peptide bonds are again not perfectly but *almost* planar—within about 4−17° (for all dihedral angles of the six atoms in the peptide bond units). The largest deviation occurs for the isomer α_L, which is in agreement with the results of ref 29 and should

TABLE 5: Relative Energies (in kcal/mol) and Dihedral Angles ϕ and ψ (in degrees) for Five Isomers of the Alanine Dipeptide (C_7^{eq}, C_5^{ext}, C_7^{ax}, β_2, α_L), Calculated with PBE, Compared with HF/MP2 and B3LYP Theory (Reference 29)

	energies			ϕ		ψ	
isomer	PBE	HF/MP2	B3LYP	PBE	B3LYP	PBE	B3LYP
C_7^{eq}	0.00	0.00	0.00	−81.7	−81.9	67.5	72.3
C_5^{ext}	1.97	1.76	1.43	−161.8	−157.3	166.1	165.3
C_7^{ax}	2.36	2.58	2.61	72.8	73.8	−55.5	−60.0
β_2	3.88	3.37	3.18	−122.7	−135.9	20.5	23.4
α_L	5.74	4.60	5.82	69.1	68.2	19.1	24.7

arise from a smaller distance and a larger interaction between the oxygen atoms of the peptide bonds.

3.3.3. Polyalanine Helix. Two stable forms of peptide helices not related by simple symmetry operations are known:[4] the 3.6_{13} (or α) helix with 3.6 residues per turn and 13 atoms in one hydrogen-bonded ring, and the 3_{10} helix (where those numbers are 3.0 and 10). The α helix dominates in most proteins; only about 3% of all residues in known proteins are in found in 3_{10} conformation. The poly*alanine* helices are the simplest realistic case for a computation. Each helix is placed along the *z* direction in an orthorhombic supercell with large lattice constants in *x* and *y* direction (30 au), periodic along the spiral with lattice constant *c* in the *z* direction.

The 3.6_{13} helix was approximated by a 3.5_{13} helix (i.e., with 3.5 residues per coil, or 7 residues per 2 coils). The periodicity length here equals two pitches (i.e., corresponds to 2 coils). The seven alanine residues contain 70 atoms in the periodically repeated supercell. For the 3_{10} helix we chose a supercell with 6 residues (60 atoms) or 2 coils. The Γ point approximation that we make in our representation of the electronic structure (see section 2) requires large supercells in all directions, and our computations satisfy this criterion.

The equilibrium geometries of both helix types are determined by separate geometry optimizations for different pitches, i.e., varying *c*. The geometry optimization of a helix takes about 1−3 dedicated hours on a Cray T3E using 64 processors (depending on the starting geometry), with about 5 min for the electronic structure calculation per geometry. Our results agree very well with the available experimental data: (i) The optimized α helix has a lower energy per residue than the 3_{10} helix, with an energy difference of 1.49 kcal/mol per residue. Hence, the α helix is more stable. (ii) The pitch of the optimized α helix is 9.78 Å, which is only about 4% lower than the experimental value (10.20 Å[4]). This decrease probably arises from the assumption that we have 3.5 residues per coil instead of 3.6, which already is 3% less. (iii) The calculated dihedral angles ϕ and ψ (of the Ramachandran plots, cf. the alanine dipeptides) of all peptide bonds agree very well with the experimental values (−60° and −40°, respectively), namely within 1.5°. Because of the large computational demand, traditional quantum chemistry (e.g., Hartree−Fock) calculations are unavailable for this system.

4. Summary

We have investigated small amino acids and peptides up to triglycine and polyalanine helices using density functional theory within a plane-wave pseudopotential scheme. The structures of a series of isomers were optimized within the local density approximation (LDA) and three different gradient corrections: BP, BLYP, and PBE. This work complements a recent density functional study of glycine[12] using a Gaussian basis set, and it is the first systematic study of alanine using density functional theory with several gradient corrections.

The alanine and glycine DFT geometries agree well with post-HF calculations and with experiment. A correct determination of the relative energies requires the inclusion of gradient corrections because the LDA greatly overestimates the strength of hydrogen bonds. For relative energies, BLYP yields the best results, but the optimal compromise between computational demand, apparent accuracy and inherent (systematic) error is PBE (or BP), because within the plane-wave pseudopotential scheme the plane-wave cutoff necessary for BLYP is excessively high (about 130 Ryd).

An accuracy of 1 kcal/mol is not attainable; however, such a resolution is often not necessary to capture the essential chemistry of the system. Force field methods lead to qualitatively incorrect results. The glycine and alanine gas-phase zwitterions are not stable and transform into the neutral isomer **II** for all forms of the DFT exchange-correlation energy, in agreement with post-HF calculations.

Finally, the structures of a few selected small peptide molecules were optimized using the PBE gradient correction. We found that all peptide bond units are planar to a high degree of precision. For the alanine dipeptide, five different conformers were considered. The equilibrium structure of the polyalanine α helix (with 70 atoms in a supercell) is in good agreement with experimental data. The α helix is more stable than the corresponding 3_{10} helix. Our study demonstrates that density functional theory is the method of choice when performing ab initio type calculations for biomolecules with $\sim$20$-$100 atoms.

Acknowledgment. The authors thank Pietro Ballone and Karl Jalkanen for helpful discussions and Silke Biermann for computational support. We are grateful to Prof. H. C. Andersen, Department of Chemistry, Stanford University, for his hospitality. This work was supported by the Deutsche Forschungsgemeinschaft (DFG).

References and Notes

(1) Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Comput. Phys. Commun.* **1995**, *91*, 215 and references therein.

(2) McCammon, J. A.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.

(3) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1.

(4) Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing, Inc.: New York, 1991.

(5) Yue, K.; Dill, K. A. *Protein Sci.* **1996**, *5*, 254.

(6) For example, see: Aida, M.; Corongiu, G.; Clementi, E. *Int. J. Quantum Chem.* **1992**, *42*, 1353.

(7) Jones, R. O.; Gunnarsson, O. *Rev. Mod. Phys.* **1989**, *61*, 689.

(8) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.

(9) Carloni, P.; Andreoni, W.; Parrinello, M. *Phys. Rev. Lett.* **1997**, *79*, 761.

(10) Montanari, B.; Jones, R. O. *Chem. Phys. Lett.* **1997**, *272*, 347.

(11) Hu, C.-H.; Shen, M.; Schaefer, H. F., III *J. Am. Chem. Soc.* **1993**, *115*, 2923.

(12) Barone, V.; Adamo, C.; Lelj, F. *J. Chem. Phys.* **1995**, *102*, 364.

(13) Sirois, S.; Proynov, E. I.; Nguyen, D. T.; Salahub, D. R. *J. Chem. Phys.* **1997**, *107*, 6770. Nguyen, D. T.; Scheiner, A. C.; Andzelm, J. W.; Sirois, S.; Salahub, D. R.; Hagler, A. T. *J. Comput. Chem.* **1997**, *18*, 1609.

(14) Csaszar, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9568.

(15) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(16) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822; *Phys. Rev. B* **1986**, *34*, 7406.

(17) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(18) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(19) Hutter, J.; Ballone, P.; Bernasconi, M.; Focher, P.; Fois, E.; Goedecker, S.; Parrinello, M.; Tuckerman, M. CPMD, Version 3.0; MPI für Festkörperforschung and IBM Research, 1990$-$97.

(20) Remler, D. K.; Madden, P. A. *Mol. Phys.* **1990**, *70*, 921; Oguchi, T.; Sasaki, T. *Prog. Theor. Phys. Suppl.* **1991**, *103*, 93. Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. *Rev. Mod. Phys.* **1992**, *64*, 1045. Galli, G.; Parrinello, M. In *Computer Simulation in Materials Science*; Meyer, M., Pontikis, V., Eds.; Plenum: New York, 1991; p 283. Hohl, D. *Theor. Chim. Acta* **1995**, *91*, 237.

(21) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.

(22) Hutter, J.; Lüthi, H. P.; Parrinello, M. *Comput. Materials Science* **1994**, *2*, 244.

(23) Fletcher, R. *Practical Methods of Optimization*; Wiley: New York, 1980; Vol. 1, p 308.

(24) Iijima, K.; Tanaka, K.; Onuma, S. *J. Mol. Struct.* **1991**, *246*, 257.

(25) Hamann, D. R. *Phys. Rev. B* **1997**, *55*, R10157 (concerning BP and PBE). Parrinello, M. *Solid State Commun.* **1997**, *102*, 107 and references therein (concerning BP and BLYP).

(26) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490, 520, 553, 587.

(27) Ding, Y.; Krogh-Jespersen, K. *Chem. Phys. Lett.* **1992**, *199*, 261.

(28) Cao, M.; Newton, S. Q.; Pranata, J.; Schäfer, L. *J. Mol. Structure (THEOCHEM)* **1995**, *332*, 251.

(29) Jalkanen, K. J.; Suhai, S. *Chem. Phys.* **1996**, *208*, 81.

(30) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L., III; Pople, J. A. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1989**, *16*, 311; *J. Am. Chem. Soc.* **1991**, *113*, 5989.